

ABSTRACT OF THE DISCLOSURE

A method for entity name and jargon term recognition and extraction. An embodiment of the present invention uses a suffix tree data structure to determine frequently occurring phrases. In one embodiment text to be analyzed is preprocessed. The text is then separated into clauses and a suffix tree is created for the text. The suffix tree is used to determine repetitious segments. Unrecognized text fragment, occurring with a high frequency, have a comparably high probability of being a name entity or jargon term. The set of repetitious segments is then filtered to obtain a set of possible entity names and jargon terms.

Patent Application No. 42390P11917